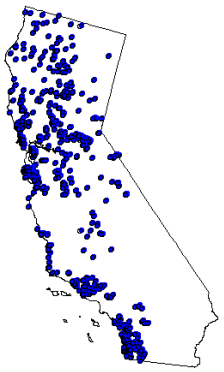
Chapter 3 STATUS OF STREAM BIOASSESSMENT ACTIVITIES IN CALIFORNIA

A few key programs in California encompass the concept and purposes of bioassessment, such that they are viable models for developing a statewide bioassessment approach. Five candidate stream bioassessment programs were identified in California based on the rigor of their scientific methods and the extent and relevancy of the data collected thus far. To qualify as a candidate program, each bioassessment program must: 1) utilize scientifically credible methods for data collection and processing, and 2) have collected a relatively large set of reliable data across a broad spatial and/or temporal scale. The following bioassessment programs in California meet these criteria: 1) California Department of Fish and Game (CDFG) Aquatic Bioassessment Laboratory (ABL) Program, 2) Lahontan Regional Water Quality Control Board Bioassessment Program, 3) U.S. Forest Service's Pacific Southwest Region Bioassessment Program, 4) U.S. Geological Survey's National Water Quality Assessment (NAWQA) Program, and 5) U.S. Environmental Protection Agency's Environmental Monitoring and Assessment Program (EMAP)/Regional Environmental Monitoring and Assessment Program (REMAP). However, it should be mentioned that the CDFG ABL provides a bioassessment support service to the state and regional boards, as well as other programs and agencies. The ABL provides sampling, taxonomic identification, and training support on a regular basis. The method developed by the ABL, the California Stream Bioassessment Procedure (CSBP) is currently the most widely used stream bioassessment method in California.

3.1 Summary of Candidate Programs

Each of the five candidate programs is summarized based on six major attributes: contact person, sampling method, timeline of sampling, data availability, purpose, and a brief description. More comprehensive summaries outlining key program elements such as habitat selection, sampling gear, sampling method, area sampled, replication, subsampling and enumeration, taxonomic identification, quality assurance procedures, data analysis/metrics, habitat assessment, and purpose for monitoring can be found in section 3.2 - Comparison of Key Elements of Candidate Programs.

3.1.1 California Department of Fish and Game (CDFG) Aquatic Bioassessment Laboratory - California Stream Bioassessment Procedure (CSBP)



The program of the California Department of Fish and Game, Aquatic Bioassessment Laboratory is designed to both investigate pollution events and to support other studies, particularly those of the RWQCBs. CDFG has been instrumental in developing technical resources and conducting numerous bioassessment studies, and in assisting with the design and collection of data for various other bioassessment programs throughout California since 1993.

Contact Person: James Harrington, State Water Quality Biologist, DFG Water Pollution Control Laboratory, 2005 Nimbus Road, Rancho Cordova, Ca 95670 (916) 358-2862 FAX (916) 985-4301 jharring@ospr.dfg.ca.gov

Sampling Method: California Stream Bioassessment Procedure (CSBP) Rapid Bioassessment Protocol (RBP)

Timeline of Sampling: 1992 - present

Data Availability: Approximately 2500 sites statewide.

Purpose of Bioassessment:

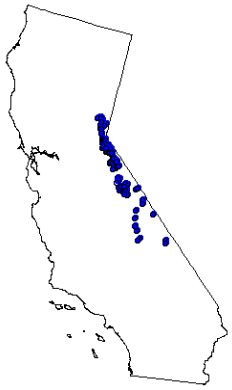
- Enforcement and resource damage assessment
- Use attainability
- Ambient monitoring
- Special studies and research

Description: DFG was the first water resource agency to be asked to assess the condition of a freshwater stream using the U.S. EPA's Rapid Bioassessment Procedure (RBPs) (Plafkin *et al.* 1989). The Lahontan Board requested the assessment in 1993 as part of the NPDES requirement of the DFG Hot Creek Hatchery in Mono County. The request necessitated the need to adapt the RBPs to California and the resulting protocol became the California Stream Bioassessment Procedure (CSBP). Because the CSBP was developed for a point-source assessment, it incorporated the use of replicated sampling of a single, richest habitat. Although not consistent with the RBP, DFG decided on this procedure for the following reasons: a) the immediate need for bioassessment was for point-source assessments, enforcements and diagnosis of known, but undocumented water quality impairment; b) there was no interest, at that time, in using bioassessment as an ambient monitoring tool; and c) the ability to produce a measure of biological metric variability at every monitoring site was deemed necessary to convince water resource managers of the robustness of biological assessments.

The CSBP is a regional adaptation of the U.S. Environmental Protection Agency (EPA) Rapid Bioassessment Protocols (Barbour *et al.* 1999). The CSBP was reviewed and refined by a CABW workgroup in 1994 and 1995 resulting in an updated version in 1996. The CSBP for wadeable streams and rivers has remained consistent over the years and is recognized by the U.S. EPA as California's standardized bioassessment procedure (Davis et al. 1996). Since 1993, the ABL has processed nearly 9000 samples collected using the CSBP at more than 2500 sites throughout California. Thousands of additional CSBP samples have been collected and processed by other entities. In addition to the CSBP for wadeable streams and rivers, as of 2002, there are versions of the CSBP for non-wadeable streams (draft), citizen monitors, lentic environments (California Lentic Bioassessment Procedure), and there is a modification of the CSBP in which samples are composited for sites that are part of an ambient bioassessment program (this CSBP modification has been adopted by the Nevada DEQ).

In addition to the numerous special studies they conduct, CDFG investigates situations where reports of activities or pollution events in the surrounding watershed may have adversely impacted stream integrity and/or stability.

3.1.2 Lahontan Regional Water Quality Control Board Biological Assessment Program – Sierra Nevada Aquatic Research Laboratory (SNARL) Method



The primary objective of this program is to incorporate consideration of biological integrity into the many regulatory and watershed management functions of the Lahontan RWQCB.

Contact Person: Thomas J. Suk, Regional Monitoring Coordinator, California Regional Water Quality Control Board, Lahontan Region, 2501 Lake Tahoe Blvd., South Lake Tahoe, CA 96150. Phone: (530) 542-5419; Email: <tsuk@rb6s.swrcb.ca.gov>

Sampling Methods: Prior to 2000, all samples were collected following protocols developed by Dr. David Herbst at the University of California's Sierra Nevada Aquatic Research Laboratory (SNARL). Starting in 2000, the Lahontan RWQCB began using and evaluating three different bioassessment sampling methods: (1) benthic macroinvertebrates, periphyton, and physical habitat assessments following SNARL protocols; (2) California Stream Bioassessment Procedures (CSBP) developed by CDFG; and (3) RIVPACS protocols being used in the Sierra Nevada by the U.S. Forest Service

Timeline of Sampling: 1995 - present

Data Availability: Approximately 350 surveys have been conducted at 200 sites in the Lahontan Region using the SNARL method. At 40 of those 200 sites, sampling was conducted using three methods (e.g., SNARL, CSBP, RIVPACS) to facilitate quantitative comparison of the results provided by each of those three methods. At approximately 30 other sites (throughout the eastern Sierra Nevada) samples were collected using both the SNARL and RIVPACS methods, and at 20 other sites (all in the Walker River drainage) samples were collected using both the SNARL and USEPA-REMAP methods. Most of this data is not yet available, and lab identification and quality assurance procedures are still underway.

Purpose of Bioassessment:

- To establish regional "reference conditions" for benthic macroinvertebrates and periphyton in streams and rivers
- To assess the impacts of human activities on the biological integrity of streams and rivers
- To evaluate the effectiveness of stream & wetland restoration efforts, BMP implementation, and permit conditions
- To develop numeric targets for TMDLs
- To develop narrative and numeric biocriteria

Description: The Lahontan RWQCB began using bioassessment in 1995, in order to monitor the success of remediation efforts at the abandoned Leviathan Mine. A more concerted (i.e., region-wide) bioassessment program was begun in 1999, for the multiple purposes outlined above.

The current regional-scale effort is focused on developing reference conditions (based on benthic macroinvertebrates and periphyton) for the eastern Sierra "ecoregion," which covers six major watershed basins (e.g., Truckee River, Tahoe Basin, Carson River, Walker River, Mono Basin, Upper Owens River). Streams in this ecoregion were stratified based on stream order, and minimally impaired sites were selected from each class of streams. Sampling has been conducted during the summer reference period (i.e., late June to early September), using protocols developed by Dr. David Herbst of the University of California's Sierra Nevada Aquatic Research Laboratory. As of this writing (i.e., 2001), the effort has focused on data collection and lab identifications; analyses of the data for biocriteria are pending. Several project-specific reports have also been generated (Upper Truckee, Leviathan, Squaw sediment TMDL)(Herbst 2002a, Herbst 2002b, Herbst 2002c).

The Lahontan RWQCB, via contract with the University of California (SNARL), is also using bioassessment data to: (1) evaluate the effectiveness of several stream & wetland restoration projects (e.g., Upper Truckee River, Bagley Valley); (2) evaluate the effectiveness of BMP implementation (e.g., Upper West Walker River, Bridgeport Valley); (3) monitor the success of remediation efforts at Leviathan Mine; (4) verify and/or assess the effectiveness of regulatory permits (e.g., fish hatcheries, Grover Hot Springs State Park); and (5) develop targets based on benthic macroinvertebrates for sediment TMDLs (e.g., Squaw Creek, Heavenly Valley Creek).

3.1.3 U.S. Forest Service - Pacific Southwest Region (California) Bioassessment Program

The focus of this program is on establishing reference conditions by collecting macroinvertebrates from a network of both perennial and intermittent wadeable streams throughout the entire state of CA, mainly on Forest Service lands. There are 18 national forests in the region (Angeles, Cleveland, Eldorado, Inyo, Klamath, Lassen, Lake Tahoe Basin Management Unit, Mendocino, Modoc, Plumas, San Bernardino, Sequoia, Shasta-Trinity, Sierra, Six Rivers, Stanislaus and Tahoe)

Contact Person: Joseph Furnish, Ecosystem Conservation Division, 1323 Club Drive, Vallejo, CA 94592

Sampling Method: Hawkins, Ostermiller, and Vinson (1998)

Timeline of Sampling: 2000 - present

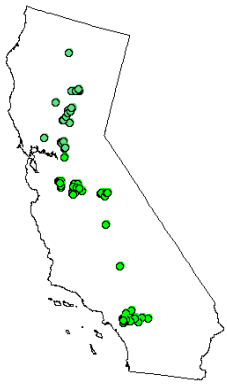
Data Availability: Approximately 176 sites in 2000 and 85 sites in 2001 located in the following watersheds: Klamath- North Coastal; Sacramento; Tulare-Buena Vista; San Joaquin; Central Lahontan; Central California Coastal; South California Coastal; North Mojave- Mono Lake.

Purpose of Bioassessment:

- Development of biocriteria and bioassessment protocol
- Monitoring of impacts from timber harvest, grazing and mining activities
- Ensure compliance with the Clean Water Act
- TMDL implementation
- Reference site characterization

Description: The primary effort has been on establishing reference conditions by collecting macroinvertebrates from a network of both perennial and intermittent wadeable streams, which can serve as the basis for monitoring biological condition and determining whether water quality has been degraded compared to reference conditions. Reference conditions will be based on development of a predictive RIVPACS (River InVertebrate Prediction And Classification System) model. Standard EPA metrics will also be considered for use if it is determined that they are sensitive to disturbances at the site and watershed (approximately 10,000-50,000 acre) scale.

3.1.4 U.S. Geological Survey: National Water Quality Assessment (NAWQA) Program



The U.S. Geological Survey (USGS) implemented the National Water-Quality Assessment (NAWQA) Program to describe the status of and trends in the quality of the nation's surface water and ground water and to provide scientific understanding of the natural and human-induced factors that affect water quality.

Contact Person: Larry Brown, Placer Hall, 6000 J St, Sacramento, CA 95819-6129

Sampling Method: USGS NAWQA

Timeline of Sampling: San Joaquin-Tulare Basins 1992-95; Sacramento Basin 1995-98; Santa Ana Basin 1998-Present.

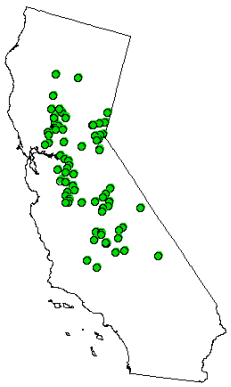
Data Availability: 17 sites in San Joaquin-Tulare Basins; 23 sites in Sacramento Basin; and 4 sites in Santa Ana Basin.

Purpose of Bioassessment:

- Describe current water-quality conditions for a large part of the Nation’s freshwater streams
- Describe how water quality is changing over time
- Improve our understanding of the primary natural and human factors affecting water quality

Description: Since 1991, the NAWQA program has been collecting and analyzing data and information in more than 50 major river basins and aquifers across the nation. The goal is to develop long-term consistent and comparable information on streams, ground water, and aquatic ecosystems to support sound management and policy decisions. Three major river basins in California were assessed as part of this program: 1) Sacramento Basin, 2) San Joaquin-Tulare Basins, and 3) Santa Ana Basin.

3.1.5 U.S. Environmental Protection Agency Central Valley Regional Environmental Monitoring and Assessment Program (REMAP)



The Central Valley REMAP project focuses on assessing the biological integrity of agriculture-dominated waterbodies located throughout California's Central Valley, which comprises more than 48,000 miles of surface water and 16 percent of the land area of California.

Contact Person: Peter Husby, USEPA Region 9 Laboratory, 1337 S. 46th St.; Bldg. 201, Richmond, CA 94804

Sampling Method: USEPA EMAP, Lazorchak and Klemm (1994)

Timeline of Sampling: 1994-1995

Data Availability: Approximately 87 sites in the Sacramento-San Joaquin Valley, covering approximately 24,000 square miles.

Purpose of Bioassessment:

- Support State of CA bioassessment and monitoring
- Assess the biotic condition of surface waters in a highly modified agriculturally influenced ecosystem.
- Determine variability of aquatic organisms in natural and man-made conveyances within the Central Valley.

Description: REMAP was initiated to test the applicability of the EMAP approach to answer questions about ecological conditions at regional and local scales. Using EMAP's statistical design and indicator concepts, REMAP conducts projects at smaller geographic scales and in shorter time frames than the national EMAP program. EMAP is a research program to develop the tools necessary to monitor and assess the status and trends of national ecological resources. EMAP's goal is to develop the scientific understanding for translating environmental monitoring data from multiple spatial and temporal scales into assessments of ecological condition and forecasts of the future risks to the sustainability of our natural resources. The objectives of REMAP are to: 1) evaluate and improve EMAP concepts for state and local use, 2) assess the applicability of EMAP indicators at differing spatial scales, and 3) demonstrate the utility of EMAP for resolving issues of importance to EPA Regions and States.

3.2 Comparison of Key Elements of Candidate Programs

A series of key elements were identified and compared among the five candidate programs. More specifically, a comparison matrix was assembled and the following elements were listed and compared: habitat selection, sampling gear, sampling method, area sampled, replication, replication as quality assurance/quality control (QA/QC), subsampling and enumeration, taxonomic level of identification, QA procedures, data analysis/metrics, and habitat assessment (Table 1). Data availability/mode of storage, written protocol availability, purpose of monitoring, and additional comments were also included but not compared in any detail as they provide very little useful information for what we are trying to accomplish in this section. Furthermore, wherever possible, the precision of each method was calculated for comparison.

3.2.1 Major Similarities and Differences Among Methods

Although all of the programs collect benthic macroinvertebrate samples to measure water quality, each has a unique goal, or question, that they are trying to address. Therefore, these differences in program goals often translate into differences in program methods. Conversely, similarities in program goals often lead to similarities in the methods. The following section briefly describes the similarities and dissimilarities of eight bioassessment method elements: habitat selection, sampling gear, collection method, area sampled, replication, subsampling and enumeration, taxonomic identification, and habitat assessment.

Habitat Selection

Most of the candidate programs focus the majority, if not all, of their sampling effort on riffle or fast-water habitats. Both CSBP and SNARL methods focus all of their sampling effort on riffle habitat. In addition to the riffle (or richest-targeted) habitat sample, USGS NAWQA also takes a separate multi-habitat sample whereby all habitats present in the reach are sampled with a proportional amount of effort going to each habitat based on occurrence in the reach. The USFS takes a similar approach in that, in addition to fast-water habitat sampling, it also collects a 10-minute qualitative sample whereby the 10-minute sampling period is apportioned so that each of the habitat types is sampled roughly in proportion to their occurrence.

The USEPA EMAP approach is slightly different from all other programs in that the amount of sampling effort is not subdivided based on habitat type, but rather the entire reach is subdivided by a number of cross-sectional transects and a sampling location is selected for each transect. Therefore, whatever habitat type is present at the selected point will be sampled. Samples collected from riffle and run habitats are composited into one sample and samples collected from pool and glide habitats are composited into another.

Table 1. Comparison of Key Elements for California Stream Bioassessment Programs

	USEPA Central Valley R-EMAP	US Forest Service	Dept. Fish & Game (CSBP)	SNARL/Lahontan	USGS (NAWQA)
Habitat Selection	• Reach determined as 40 times the wetted width with a minimum reach length of 150 meters and maximum length of 500 meters.	• Fast-water (Almost always riffles, runs may also be sampled), four consecutive areas within the sample reach. • Reach length may vary from about 200-500 meters	• Stream reach selected which contains at least 5 riffles within the same order and relative gradient. • If no riffles are present, or less than five within a reasonable distance, EMAP selection method is used as default.	Riffles within 150 m study reach.	• All habitats in selected reach (QMH sample) • "Richest-targeted sample" (RTH sample) with riffles being the priority habitat and woody debris sampled when riffles not available.
Sampling Gear	Rectangular net 50 cm wide, 500 µm mesh.	Surber sampler (0.09 m^2), 500 µm mesh, 1-meter long net to prevent backwashing	30 cm wide D-shaped kick net (500 µm mesh)	30 cm wide D-frame net (250 µm mesh)	• RTH: .5 m x .25 m net with 425 µm mesh. • QMH: standard d-frame net with 210 µm mesh.
Collection Method	• Samples collected at 9 evenly spaced transects within reach. • Composited as riffle/run or glide/pool, 0-9 samples per composite.	• Fixed area sample is composed of 8 Surber samples (4 riffles x 2 samples from each riffle) • 10-minute qualitative sample from all major habitats present.	One composite of 3 samples is collected from the upstream third of 3 randomly chosen riffles.	Each sample is a composite of 3 samples taken from each of 5 randomly selected riffles.	• RTH: composite of samples from 5 locations within riffles. . • QMH: equal effort in all habitats present in entire reach. Time variable (usually 1 hr).
Area Sampled	• Area per sample is ~ 0.5 m^2 • Area per composite is variable depending on proportion of habitat type sampled.	• Total area sampled per fixed area composite = 0.72 m^2 • Total area sampled for fixed time samples variable.	Total area sampled per composite = 0.54 m^2 Total area sampled per site = 1.62 m^2	Total area sampled per composite = 0.27 m^2 Total area sampled per site = 1.28 m^2	• RTH: Total area sampled per composite = 1.25 m^2 • QMH: Total area sampled variable
Replication	• No site replicates.	• No site replication using the same methods.	3 randomly-selected samples taken at each site	5 randomly-selected riffles from each site	Limited replication.
Replication as QA/QC	• Same season, different team revisits (2 sites) • Next year revisits (10 sites).				• Replication limited to a subset of 4-6 sites • 3 samples are collected at each site

Table 1. Comparison of Key Elements for California Stream Bioassessment Programs (continued)

	USEPA Central Valley R-EMAP	US Forest Service	Dept. Fish & Game (CSBP)	SNARL/Lahontan	USGS (NAWQA)
Subsampling And Enumeration	• Random subsampling to 300 organism count/identification	• Composite samples are divided into equal-sized proportions and all organisms are removed from each sub-sample until a minimum of 500 specimens (early data was 300) have been obtained from a complete sort of 1 or more subsamples. • Big/rare specimens are also removed from the entire remaining sample during a 10-minute examination.	• 300 organisms for ID. • All organisms in grid are counted for abundance	• Subsampling using rotating drum splitter • Minimum count of entire split = 250 organisms, (actual range = 300-500) • Big/rare organisms are also removed	• Field splits conducted when sample volume is >0.75 L. • Field processing can result in 4 sample components: large-rare, main-body, elutriate, and split-sample. • Samples are split until composite volume is ≤ 0.75 L.
Taxonomic Level of ID	• Lowest taxon possible • Genus, species, or species group (including Chironomids and Mites).	• Insects are primarily identified to the genus level. • Chironomidae are identified to the sub-family level. • Non-insect invertebrates identified to various levels depending on available keys.	• Insects are primarily identified to the genus level. • Chironomidae are identified to the sub-family level. • Non-insect invertebrates identified to various levels depending on available keys.	• Lowest taxon possible • Genus, species, or species group (including Chironomids and Mites).	• Most insects to species or genus. • Other organisms variable.
QA Procedures	• Field: revisit by different team - same year (2 sites) and second year revisit on 10 sites • Vouchers and reference collection maintained • Lab: sorting checks 10%; ID checks 100%.	• Field: instrument calibration. • National Aquatic Monitoring Center (NAMC) procedures for sample processing. • Vouchers and reference collection maintained at NAMC.	• Field: crew members trained for sampling consistency, and audits • Lab: sorting checks 100%; ID checks 10-20%, bioassessment validation 10-20% • Internal and external QC, 10% each	• Field: instrument calibration, crew training. • Vouchers and reference collection maintained • Lab: sorting checks 20%; ID checks 100%. • Lab training and corrective actions.	• All identifications by qualified experts • 10 % internal QC • External vouchers
Data Analysis/ Metrics	Various including many alternatives for use in screening environmental correlation.	No standard procedure has been designated. RIVPACS will be utilized to develop a model to determine the level of impact to the biological assemblage at a site. Benthic-IBI may also be used depending on performance.	Developed own multimetric and multivariate approach.	Various including many alternatives for use in screening and environmental correlation.	No established metrics or endpoints used. Analysis emphasizes multivariate gradient analyses.
Habitat Assessment	Quantitative surveys of 11 transects (intensive) and full reach (water and sediment chemistry, thalweg, width, depth, velocity, substrate, etc.	1) Densiometer shade measurements, 2) wetted width, 3) mean depth (n=3 measures x 10 transects= 30), 4) substrate- Wolman pebble count, 5) conductivity, 6) alkalinity, 7) Gradient, 8) Habitat Types (Montgomery-Buffington channel classes)	EPA method and additional: • Canopy • Quantitative substrate • Pebble count • Substrate consolidation • Depth & width • Velocity	Quantitative surveys of 15 transects (intensive) and full reach (chemistry, width, depth, velocity, substrate, etc.)	Detailed habitat measurements at various scales (basin, segment, reach, transect). Protocols now call for 11 habitat transects within each reach.

Table 1. Comparison of Key Elements for California Stream Bioassessment Programs (continued)

	USEPA Central Valley R-EMAP	US Forest Service	Dept. Fish & Game (CSBP)	SNARL/Lahontan	USGS (NAWQA)
Data Availability and Mode of Storage	Obtained Excel spreadsheets for Central Valley 1994 & 1995 macroinvertebrate data (no habitat data)	Data are available from the NAMC and eventually will be deposited into the USFS corporate database system of the Natural Resource Information System (NRIS).	Access database (Cal EDAS). Much data still in Excel.	Obtained 4 Excel spreadsheets: Upper Truckee River 1998-2000, Leviathan Mine Watershed 1999, Leviathan Spring 1995/1997, Leviathan Fall 1998)	Obtained Excel spreadsheets for Sacramento River Basin 1996-1998 invertebrate data (no habitat data)
Written Protocols Availability	Lazorchak and Klemm, 1994.	Hawkins et al. 1998	Yes	http://www.swrcb.ca.gov/rwqcb6/QAPP/QAPP_Index.htm	http://water.usgs.gov/nawqa/protocols/doc_list.html
Comments		• Analysis tools not fixed, intend to use both multimetric and multivariate approaches. • Approximately 170 prospective reference sites sampled during FY2000 to develop a RIVPACS model.	• Calibration with RIVPACS and EMAP • More than 8000 samples to date	• Calibration with CSBP & RIVPACS underway. • Analysis tools not fixed, intend to use both multimetric and multivariate approaches. • Approximately 225-250 streams sampled to date (1996-2000). • About 25-50 of these are monitored annually or even seasonally.	Program is in support of the National Water Quality Assessment Program and does not include continuous (annual sampling). Intensive sampling typically only occurs for a year or two.
Purpose for Monitoring	• Support State of California bioassessment and monitoring. • Assess the biotic condition of surface waters in a highly modified agriculturally influenced ecosystem. • Determine variability of aquatic organisms in natural and man-made conveyances within the Central Valley.	• Development of biocriteria & bioassessment protocol • Monitoring of impacts from timber harvest, grazing and mining activities • Ensure compliance with the Clean Water Act • TMDL implementation	• Enforcement and resource damage assessment • Use attainability • Ambient monitoring • Special studies and research • Develop and promote bioassessment methodologies • Test and troubleshoot methods	• Biocriteria development and assessment & monitoring. • Livestock grazing stream restoration • Acid Mine Drainage stream restoration monitoring. • TMDL development for sediments. • Reference condition sampling	In support of National Water Quality Assessment Program, a water quality program. Biological assessments are included as a measure of ecological health of streams.

Habitat Selection

Most of the candidate programs focus the majority, if not all, of their sampling effort on riffle or fast-water habitats. Both CSBP and SNARL methods focus all of their sampling effort on riffle habitat. In addition to the riffle (or richest-targeted) habitat sample, USGS NAWQA also takes a separate multi-habitat sample whereby all habitats present in the reach are sampled with a proportional amount of effort going to each habitat based on occurrence in the reach. The USFS takes a similar approach in that, in addition to fast-water habitat sampling, it also collects a 10-minute qualitative sample whereby the 10-minute sampling period is apportioned so that each of the habitat types is sampled roughly in proportion to their occurrence.

The USEPA EMAP approach is slightly different from all other programs in that the amount of sampling effort is not subdivided based on habitat type, but rather the entire reach is subdivided by a number of cross-sectional transects and a sampling location is selected for each transect. Therefore, whatever habitat type is present at the selected point will be sampled. Samples collected from riffle and run habitats are composited into one sample and samples collected from pool and glide habitats are composited into another.

Sampling Gear

The majority of candidate programs prefer to use D-frame or rectangle frame kicknets to collect samples; however, net mesh size is variable among programs. Most of the methods prefer a net with a mesh size around 500 μm. For example, both CSBP and USFS methods use 500 μm mesh netting, while USEPA EMAP and USGS NAWQA (RTH sampling) use 595/600 μm and 425 μm, respectively. On the other hand, SNARL prefers 250 μm mesh netting, and USGS NAWQA (QMH sampling) uses 210 μm mesh netting.

The only obvious difference in sampling gear, other than mesh size, is USFS method's use of a Surber sampler. All other programs use either a D-frame net or rectangle frame kicknet to collect samples. CSBP, SNARL, and NAWQA (QMH) methods all use D-frame nets. Both EMAP and NAWQA (RTH) methods use rectangle frame kicknets.

Collection Method

Perhaps the largest difference between programs lies in the collection method used by each. All of the programs take one or more composite samples from each site, but the make up of and method of collecting each composite is quite variable. For a detailed description of each programs' sampling method see Appendix B.

Area Sampled

The area sampled per composite is quite variable ranging from 0.27 m^2 for the SNARL method to 1.25 m^2 for NAWQA (RTH) method. However, composites using the EMAP method may sample up to 4.5 m^2, but the area sampled varies based on habitat selection. The total area sampled per reach, not including fixed time or QMH sampling, ranges from 0.72 m^2 for the USFS method to 4.5 m^2 for the EMAP method.

Replication

Only three of the five methods collect valid site replicates as part of their sampling programs. Both the CSBP and SNARL methods routinely collect replicate samples at every site (i.e., three and five, respectively), whereas NAWQA collects replicate samples at a subset of 4-6 sites per study. USFS collects no replicates samples, and EMAP only collects QA/QC replicates using same season, different team revisits and same team, different year revisits.

Subsampling and Enumeration

Both the count and method of subsampling is highly variable among all programs. NAWQA uses both a qualitative visual sort method and a quantitative fixed-count method of subsampling; however, the organism count varies based on the data quality objectives of the study. Both the CSBP method and the EMAP method subsample to 300 organisms, but the remaining programs use subsampling methods based on composite sample splits and identifying the entire split to within a range of organisms. For example, USFS divides the composite into equal-sized portions and all organisms are removed until a minimum of 500 specimens have been obtained from a complete sort of one or more subsamples. The SNARL method uses a similar subsampling strategy whereby the composite sample is split until the minimum count of the entire split is 250 organisms.

Taxonomic Identification

Most of the programs identify insects to the lowest taxon possible, which is usually the genus and/or species level. However, USFS and CSBP identify Chironomid midges to the sub-family level. Non-insect invertebrate identification is variable, usually depending upon available taxonomic keys.

Habitat Assessment

Habitat assessment tends to be highly variable among programs in terms of rigor and detail of measurements. EMAP, NAWQA, and SNARL collect quantitative measurements at multiple (11-15) transects throughout the study reach, utilizing a relatively comprehensive habitat assessment approach. On the other hand, CSBP and USFS utilize more rapid habitat assessment techniques (visual-based for most measures) to characterize physical habitat semi-quantitatively.

3.2.2 Comparison of Performance Characteristics for Bioassessment Methods

Although water quality programs have distinct goals for conducting bioassessments and require different levels of effort in sample collection, taxonomic identification, and data analysis, discrete methods may yield comparable data for certain objectives despite these differences in effort. If discrete methods are similar with respect to the quality of data they produce, it is possible to use the results together. In other words, determining the performance characteristics of individual methods enables agencies to share the results of bioassessments by providing an estimate of the level of confidence in assessments from one method to the next (Barbour et al. 1999). The best way to determine the quality of data produced by a method is through the use of

data quality objectives. Data quality objectives (DQOs) are qualitative and quantitative expressions that define requirements for data precision, bias, method sensitivity, and range of conditions over which a method yields satisfactory data (Klemm et al. 1990).

The documentation of performance characteristics for all methods is known as the performance-based method system (PBMS – see ITFM 1995), which is essentially a system that permits the use of any method of sampling and analysis that meets established requirements for DQOs (Diamond et al. 1996, NWQMC 2001). The basic elements of a PBMS approach include method precision (repeatability of measurements), bias (skewness of measurements), sensitivity (detection limit), and accuracy (proximity to the analytical truth).

For the PBMS approach to be useful, three basic assumptions must be met (ITFM 1995):

1. DQOs must be set that realistically define and measure the quality of the data needed; reference (validated) methods must be made available that meet those DQOs;
2. there must be proof that the method yields reproducible results that are sensitive enough for the program; and
3. the method must be effective over the prescribed range of conditions in which it is to be used.

Key Performance Characteristics
- Precision
- Sensitivity

For bioassessments, the above assumptions imply that a given method for sample collection and analysis produces data of known quality, including precision, the range of habitats over which the collection method yields a specified precision, and the magnitude of difference in data among sites with different levels or types of impairment (Diamond et al. 1996). Calculating the performance characteristics for a given bioassessment method is essential to understanding the robustness of the method for reliably determining the condition of the aquatic ecosystem. A method that is very labor intensive and requires a great deal of specialized expertise, and, in turn provides a substantial amount of information, is not necessarily the most appropriate if it is not very precise and repeatable. A less rigorous method may be less sensitive to detecting perturbation or have more uncertainty in its assessment. All of these attributes are important to minimizing Type I and II error in bioassessment. The ultimate question resides in a firm balance between cost and resolution, i.e., is more information better (more cost) or is a limited amount of the right information best (less cost). A knowledge of method precision, sensitivity, bias, and accuracy helps with this decision. For purposes of this discussion, the key performance characteristics are precision and sensitivity to establish a basis for understanding the CSBP and SNARL methods comparison presented later in his section.

Establishing DQOs for a bioassessment method helps to evaluate the adequacy and robustness of a method. For example, we may establish the following DQOs:

DQO 1. We want to be able to detect a 20% change, e.g., five categories of condition on a 100-pt scale for a calibrated biological index.

DQO 2. We want the method to have a discrimination efficiency of greater than 75%, i.e., the method is calibrated so that only 25% or less ($\beta = 25\%$) of the *a priori* determined sites of reference and degraded would be misclassified.

Using these two example DQOs, we establish the following hypothetical scenario.

Hypothetical Scenario

To conduct an analysis of the performance of a bioassessment method, or several methods, five steps can be identified: 1) compare the relative variability of the various methods from both reference and degraded sites – *DQO 1*, 2) evaluate sensitivity or discrimination efficiency – *DQO 2*), 3) evaluate precision, 4) evaluate bias and accuracy, and 5) evaluate ability to make a correct assessment – *DQO 2*. In this hypothetical example, we compare three methods used side-by-side to collect bioassessment data.

Step 1 (Characterization of sites). The first step toward evaluating a method's performance as a bioassessment tool, is to collect or assemble data from both reference and degraded sites. Having a population of reference sites as well as a population of data collected from known degraded sites is essential for determining both the relative performance using different levels of biological condition as well as determining sensitivity or discrimination efficiency. Box-and-whisker plots are used to plot data for a given biological indicator (e.g., a metric or index) from each of the three methods (Figure 1). These plots illustrate the amount of variability measured in a population of sites (in both reference and degraded categories). For this example, we will say that methods 1 and 2 have tight enough ranges in variability to allow us to meet the first DQO, i.e., an ability to detect a 20% change.

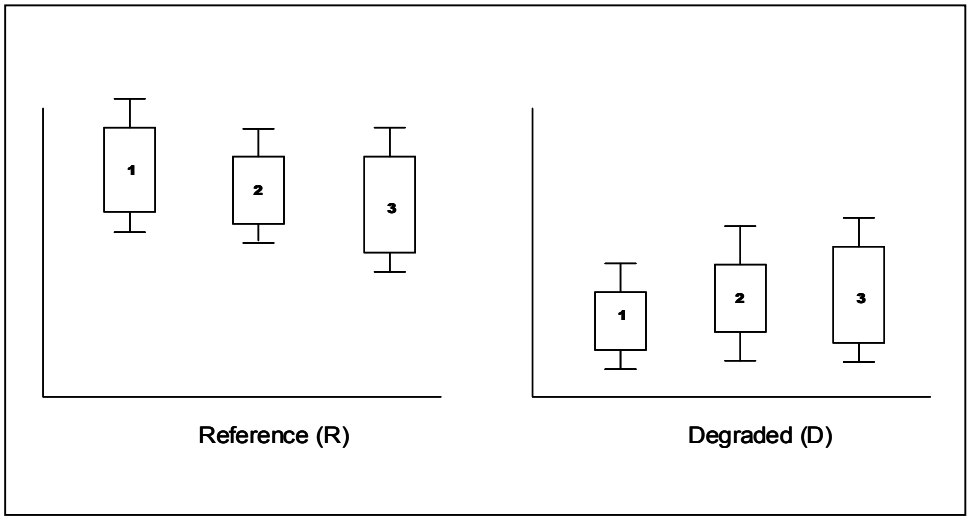


Figure 1. Box-and-whisker plots showing the distribution of data collected from reference and degraded sites using three separate methods (1, 2, and 3). Boxes illustrate population attributes (via percentile distribution, i.e., 25% - 75%) and whiskers provide a sense of variability.

Step 2 (Sensitivity). The second step is to evaluate the sensitivity of each method, or ability to discriminate between reference and degraded sites. By examining the reference and degraded box and whisker plots side-by-side, it is possible to determine the sensitivity of a given method. The reference and degraded plots are paired to show the amount of overlap, or lack thereof (Figure 2). The more overlap between plots the less sensitive the method, and vice versa. In this example, method one is the most sensitive because there is no overlap between plots, and method three is the least sensitive because it has the most overlap of the interquartile ranges. Method 1 meets the second DQO of having greater than 75% discrimination efficiency.

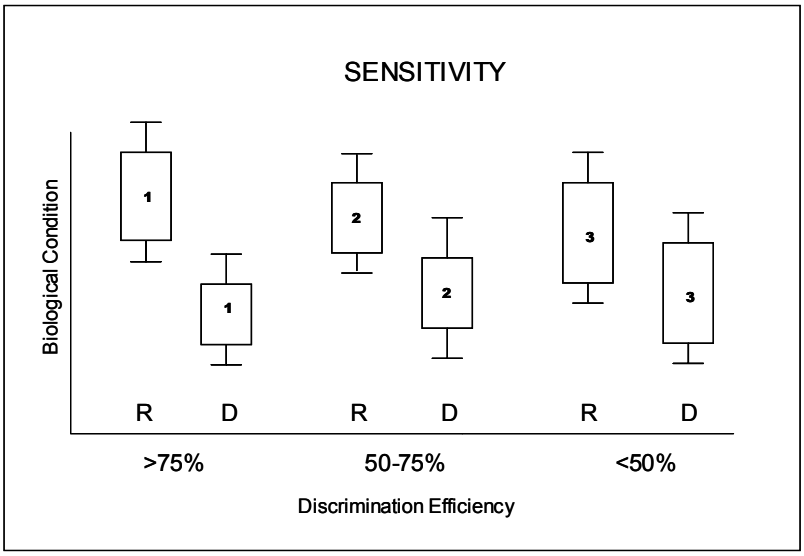


Figure 2. Box-and-whisker plots illustrating the ability of each method to discriminate between reference (R) and degraded (D) conditions. Method one discriminates greater than 75% of the sites correctly; method two can only discriminate between 50 and 75% of the sites correctly; and method three is least sensitive, discriminating less than 50% of the sites.

Step 3 (Precision). The third step is to evaluate the method precision, or repeatability of measurements, using all sites (i.e., reference and degraded) in the population. Repeated samples (replicates or duplicates) are required to calculate the standard deviation from the mean. This can be illustrated by graphing the mean value for a given metric or index and incorporating error bars to show the standard deviation (Figure 3). In this example, method two is the most precise because it has the smallest standard deviation around the central tendency (mean), and method three is the least precise because it has the largest deviation around the mean.

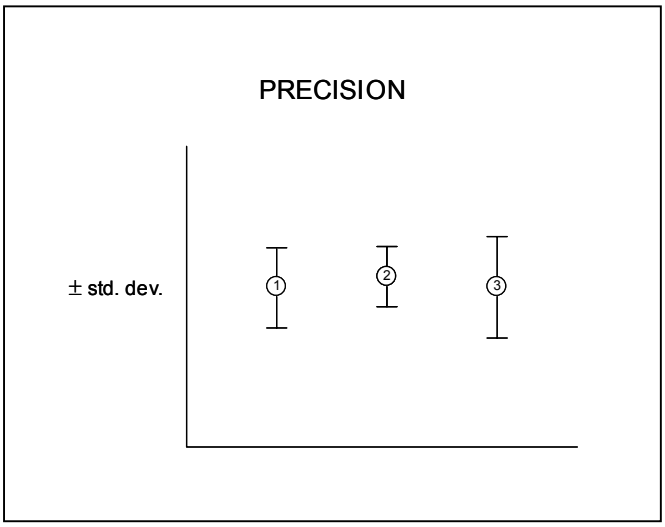


Figure 3. Graph illustrating the precision of each method for a given measure using means and standard deviations.

Step 4 (Bias and Accuracy). Although not treated here, bias and accuracy are often determined for various components of bioassessment, such as laboratory subsampling and taxonomic identification. In the laboratory setting, it is relatively easy to determine the accuracy of sorting as well as the bias of sorters and taxonomists through the implementation of simple QA/QC plans. For example, after organisms are identified, they can be sent to another independent taxonomist for confirmation of taxonomic identifications. Bias would be a consistent mis-identification that could be ascertained through QC checks. Additionally, after a sample is sorted, an assigned QC officer can resort the sample to determine the percentage of "missed" specimens. Bias might be in always missing midges, or very small specimens, for example. While both bias and accuracy can be determined at various stages in the bioassessment process, it is often unclear how these characteristics can be calculated for the overall assessment where "truth" is determined by an impairment threshold.

Step 5 (Site assessment). The fifth and final step is to evaluate the influence of the performance characteristics on making a correct assessment. By examining the performance characteristics of the three methods in relation to a fixed impairment threshold, we can determine a level of confidence in each index value (Figure 4.) In this example, we use the three methods at one site and their measurement precision and discriminatory efficiency to illustrate how a site assessed as impaired by all three might be evaluated. For Method 1, we have high discrimination efficiency and moderate precision. Because the value of the site and its error bars (precision) fall below the impairment threshold, we have a high level of confidence that this site is in fact impaired. Method 3 is the least precise and least discriminatory, and thus, our confidence that this site is impaired is low. For Method 2, which has the highest precision, the site would likely be assessed as impaired. However, the discrimination efficiency of Method 2 indicates that we only assess between 50 and 75% of our sites correctly. In this case, sites that are slightly

impaired, i.e., near the threshold, would benefit from additional, supplemental data (e.g., complementary water or habitat quality data, a follow-up biosurvey, etc.).

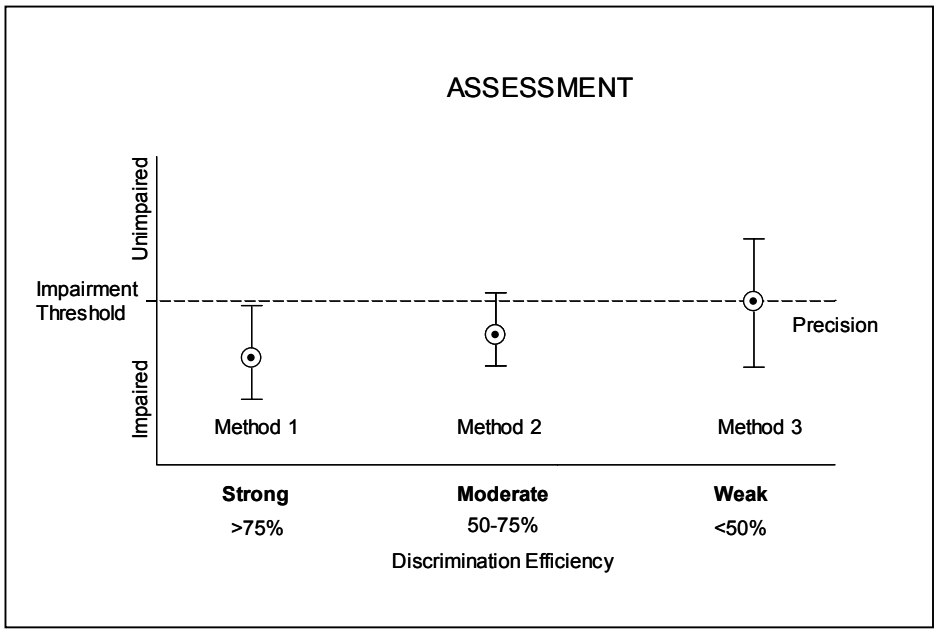


Figure 4. Graph illustrating the ability of a method to yield a correct assessment based on a combination of precision and sensitivity (or discriminatory efficiency) and the value of the assessed site in relation to the impairment threshold.

Comparison of CSBP and SNARL Methods

Due to the paucity of data provided to us at the time of this report, only one performance characteristic, method precision (i.e., measurement error within a site), could be evaluated for two candidate methods, CSBP and SNARL. It should be mentioned, however, that there are a few caveats with this precision comparison. First, the populations sampled using each method were quite different from each other. The SNARL method sampled primarily high elevation streams (5,000-7,500 feet) in the Sierra Nevada Mountain Range, whereas the CSBP collected samples across a wide variety of locations and across multiple ecoregions, primarily in lower elevation streams. Because variability is a combination of both natural variability and measurement error, greater variability does not necessarily imply greater measurement error when two distinct populations are sampled. Consequently, a side-by-side comparison would help to minimize the influence of natural variability and allow a more accurate comparison of measurement error between these two methods. Secondly, the net mesh size used in the SNARL method and CSBP is very different, 250 μm and 500 μm respectively. This difference can introduce a good deal of variability in the results because of organism selectivity (bias) associated with each method. However, it is uncertain as to whether this would significantly affect the comparison of precision estimates and requires further research. Thirdly, it is uncertain what types of sites (i.e., impacted, reference, etc.) and in what proportions these types of sites make up the datasets that were analyzed. Different types of sites may introduce more

natural variability among replicates than others, and thus, could affect the precision estimate for that method. With this simple comparison, we provided estimates that the SNARL method may be more precise, except for the caveats cited previously. We do not know if the higher precision is either ecologically or statistically significant, and if so, whether cost implications justify the increased precision. However, this exercise demonstrates one of the steps necessary for adequately comparing methods.

As a focus of this methods comparison, sampling precision was evaluated using the root mean square error (RMSE) to measure variability. RMSE, also called the standard error of estimate, is an estimate of the standard deviation of a population of observations. The RMSE was calculated for eight common biological metrics used by both the CSBP and SNARL methods. RMSEs ranged from 0.72 to 11.78 for CSBP and from 1.03 to 7.78 for SNARL for the eight metrics (Table 2). The RMSE was lower for CSBP than for SNARL for the richness metrics (i.e., total number of taxa, EPT taxa, and components of the EPT – Ephemeroptera, Plecoptera, and Trichoptera). However, the reverse was true for the composition and tolerance metrics (i.e., %EPT, %Tolerant organisms, and %Dominance). The relative spread of the values for the two methods is illustrated when the mean and standard deviation for each metric are graphed (Figure 5). The SNARL method recorded a higher mean for each metric. However, the standard deviation was generally lower for the CSBP method.

Table 2. Comparison of ANOVA results between CSBP and SNARL methods.

Metric	*CSBP*			**SNARL**			*RPD RMSE*	*Difference CV*
	RMSE	*MEAN*	*CV*	*RMSE*	*MEAN*	*CV*		
Total Number of Taxa	3.21	16.72	19.23	3.76	27.09	13.9	15	5.4
EPT Taxa	1.59	6.45	24.71	1.85	11.1	16.67	15	8.04
Ephemeroptera Taxa	0.72	2.97	24.44	1.03	6.77	15.26	35	9.18
Plecoptera Taxa	1.09	2.83	38.54	1.26	4.33	28.99	10	9.55
Trichoptera Taxa	1	2.82	35.65	1.16	5.73	20.22	15	15.43
%EPT	11.29	42.21	26.76	9.5	63.32	15	17	11.76
% Tolerant Organisms	11.24	22.37	50.23	5.4	11.32	47.7	70	2.53
%Dominance	11.78	43.45	27.12	7.78	36.16	21.52	41	5.6

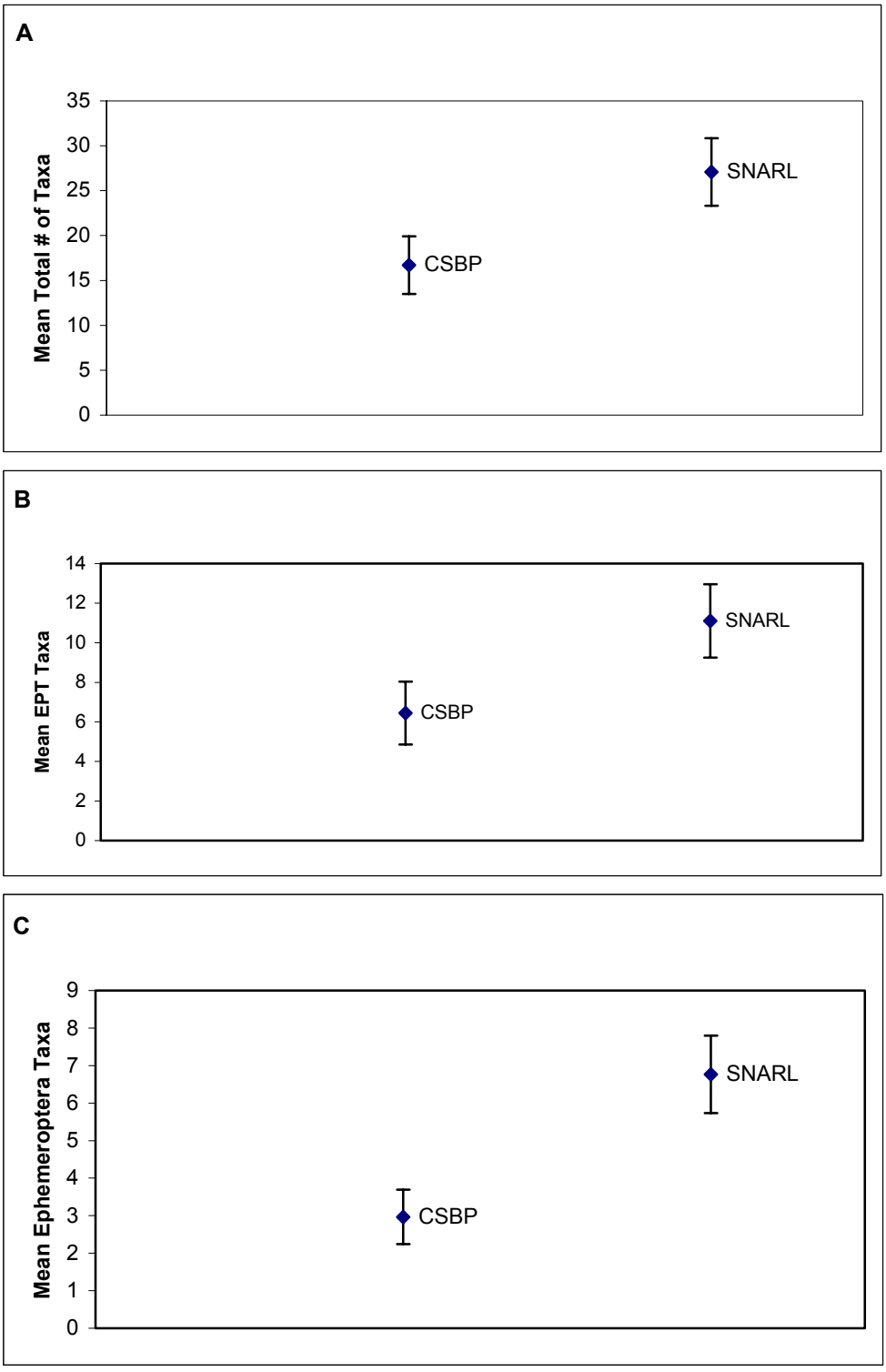


Figure 5. Comparison of precision (mean ± 1 s.d.) between the CSBP and SNARL methods for representative biological metrics for richness (graphs a-e), composition (f-g), and tolerance (g-h).

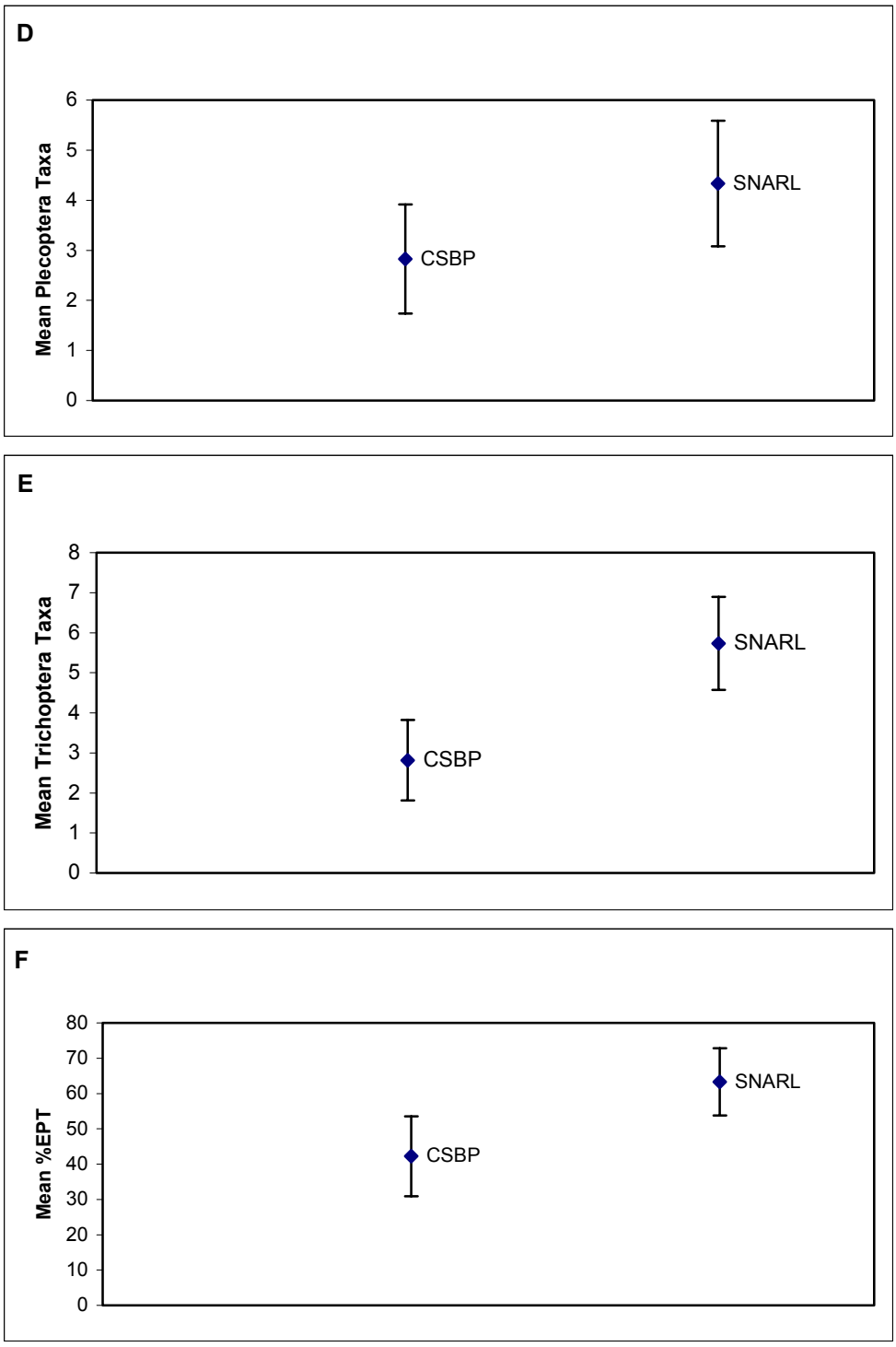


Figure 5 (continued). Comparison of precision (mean ± 1 s.d.) between the CSBP and SNARL methods for representative biological metrics for richness (graphs a-e), composition (f-g), and tolerance (g-h).

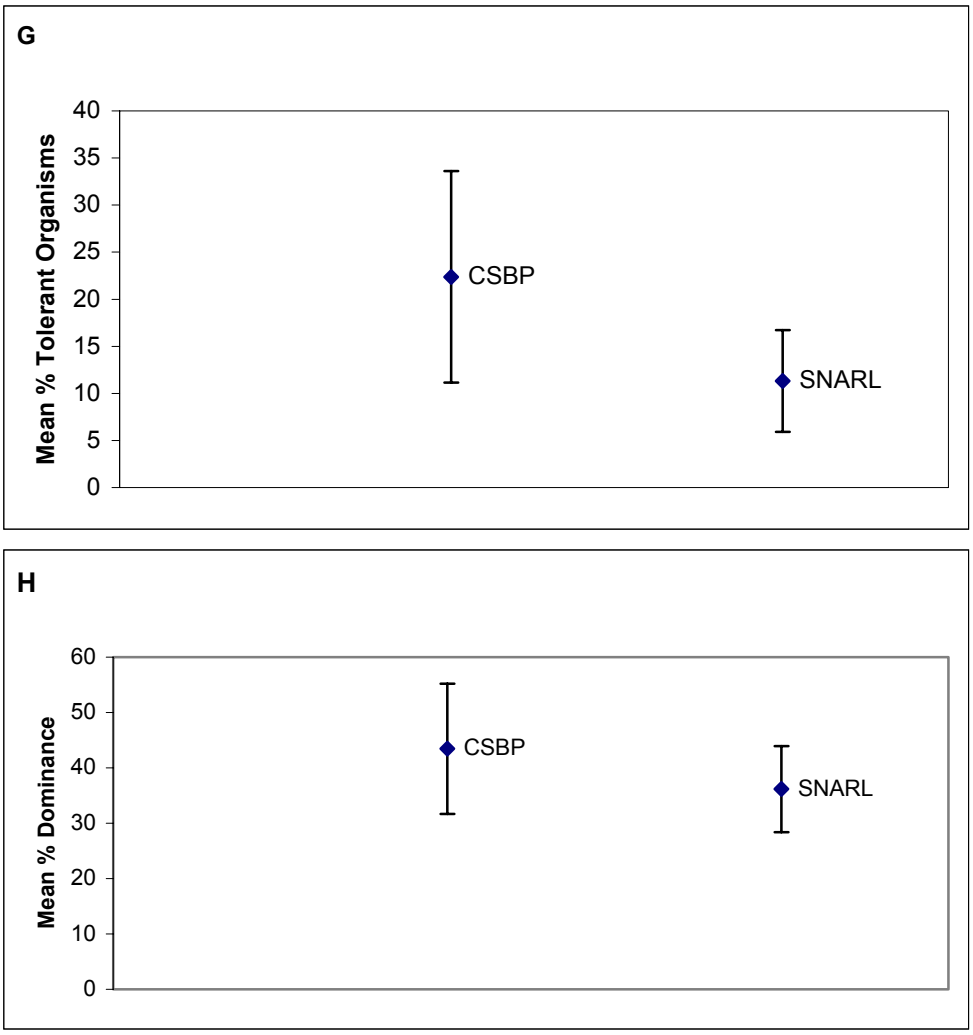


Figure 5. Comparison of precision (mean ± 1 s.d.) between the CSBP and SNARL methods for representative biological metrics for richness (graphs a-e), composition (f-g), and tolerance (g-h).

Because various components of these methods were vastly different, the coefficient of variation (CV) was calculated to evaluate the variation adjusted for the mean of each metric. The values of the CV were lower for the SNARL method for all eight metrics. However, because there are no calibrated indexes and impairment thresholds established for these methods, we do not know whether the lower CVs for the SNARL are ecologically significant. As a point of discussion, we can draw from our DQO 1 established as part of our hypothetical example. Although the difference in the CV values between the two methods never exceeded 20%, the majority of the individual metrics for each method did exceed 20% (our initial DQO from the hypothetical example). It should be noted that our DQO 1 is established for a calibrated index and not individual metrics. However, the precision for overall index scores are often more precise than for individual metrics (Stribling et al., in review). For example, Stribling et al. found that for

three separate data sets (Maryland DNR, Prince George's County DER, Wyoming DEQ), the overall index score was consistently more precise than for any of the individual metrics, with one exception. Still, overall index precision cannot be easily speculated given the precision of only a few individual metrics. One critical step would be to develop a biological index for each method, and then compare the overall index precision to get a better understanding of which method is more precise. Depending on the outcome, another critical step would be to calculate a power cost efficiency (PCE) analysis (Barbour and Gerritsen 1996) to evaluate the cost implications of the added precision that might be realized from a more rigorous method.

Conclusions

From this simple comparison study with an incomplete data set, the results are inconclusive about the performance of the CSBP method vis-à-vis the SNARL method, and vice versa. However, Dr. David Herbst of the University of California Sierra Nevada Aquatic Research Laboratory has conducted a side-by-side comparison of these two methods along with a third method, USFS, also referred to as RIVPACS. Data analyses are ongoing and the results should be available near the beginning of 2003 (Herbst and Silldorff 2003). Furthermore, CDFG-ABL is currently conducting a side-by-side comparison of the CSBP, RIVPACS, and USEPA EMAP methods using a slightly larger dataset (approximately 240 sites from all over the state). This study is ongoing and the results are not yet available. We recommend that the results of these comparisons be sought and considered by anyone who is interested in the performance characteristics of these methods. In order to foster a valid scientific comparison of the performance and cost-effectiveness of a method, or multiple methods, several pieces of information must be made available:

- a data set of both known degraded and qualified reference conditions
- repeated samples (replicates or duplicates) to calculate the standard deviation from the mean (from both degraded and reference sites)
- DQOs from the QA/QC plan
- costs associated with the different levels of subsampling (for cost efficiency calculations)
- number of subsamples required to detect differences in the data
- discrimination (i.e., power) that is required to detect differences in the data.

A case example of how the Florida Department of Environmental Protection (DEP) examined the performance characteristics of their collection and assessment methods can be found in Appendix C.

3.3 Integrating Disparate Programs

The integration of discrete programs is primarily dependent on the results of the performance characteristic characterization. If it is evident that the quality of data is comparable among programs, then it is possible to integrate results of assessments among programs. Essentially, it is the quality and detail of data that defines the level of integration of disparate programs.

However, there are several elements that widely differ among the programs and may hinder the integration of actual biological data:

- Mesh size that retains/excludes certain organisms
- Level of subsampling & enumeration
- Sampling area and method
- Taxonomic resolution

Although there is a certain amount of disparity among all the candidate programs in each of these elements, most will likely allow a certain level of integration provided that the DQOs yield comparable data. This could ultimately lead to an integrated set of reference sites, which could be used to characterize reference conditions all throughout California. The features or attributes proffered by these candidate programs for integrating ecological information include:

- Candidate reference sites
- Identification of impaired sites or sites at risk
- Characterization of watersheds and stream reaches
- Quality ratings for water resource management
- Taxonomic distribution list and statewide records

3.4 Recent Initiatives in Bioassessment

A few recent and notable bioassessment initiatives in California include the development of 1) an Inter-laboratory Quality Assurance/Quality Control (QA/QC) Program, 2) the CalEDAS Database, 3) an Index of Biological Integrity (IBI), and 4) a standardized methodology of reference site selection for wadeable streams.

3.4.1 Inter-laboratory Quality Assurance/ Quality Control Program

Bioassessment data are being collected in California at a rapidly increasing rate. Since there will be much more taxonomic identification work than can be managed by a single laboratory, the standardization of laboratory techniques and taxonomic data is critical to sharing data analyzed by different laboratories.

In 1999, DFG-ABL instituted an inter-laboratory quality assurance/quality control (QA/QC) program for taxonomic identification. There are two main goals of an external QA/QC program, 1) to assess the quality of taxonomic data and its impacts on bioassessment metrics and 2) to assure that taxonomic data from different sources can be included in a common database. The QA/QC procedures are designed to help ensure compatibility of data among different macroinvertebrate laboratories and to ensure taxonomic consistency and high quality of taxonomy for all laboratories involved.

The DFG QA/QC procedure compares each taxonomic identification and groups of all discrepancies into two categories, 1) identification discrepancies, and 2) relative taxonomic effort discrepancies. Identification discrepancies are instances in which the two laboratories do

not agree on the identification of a particular taxon. Relative taxonomic effort discrepancies are cases in which the original taxonomic determination is less or more precise than that of the QC laboratory. Although these differences in taxonomic effort are not as obvious as disagreements over identification, they can have a very strong impact on metrics calculations and often make up the majority of differences in the taxa lists of different laboratories. In addition to taxonomic discrepancies, the procedure evaluates differences in enumeration by the two laboratories. Small differences are a common occurrence in QC analysis and should not be a cause for concern unless the discrepancies are large.

The current external QA/QC program only involves assessment of taxonomy and enumeration; it does not include checks of subsampling procedures. A QA/QC protocol for sub-samples may be included in future programs, but at this point, it is considered the internal responsibility of each laboratory.

3.4.2 CalEDAS Database Development

As bioassessment has become increasingly more included in California's water quality management programs, the amount of biological community data and associated physical and chemical data collected around the state has grown at a rapid pace. The benefits of being able to manage and manipulate this data in a consistent way are immense; these data will ultimately provide the basis for fully exploiting bioassessment's potential as a water quality management tool.

Since 1998, DFG-ABL has been developing a Microsoft Access® database for managing its own bioassessment datasets. CalEDAS is a modification of the EDAS® (Environmental Data Analysis System), which was developed by Tetra Tech, Inc. for the USEPA. The main taxonomic table in CalEDAS (the Benthic Master Taxa List) is based on the CAMLnet List of Standard Taxonomic Effort. DFG-ABL uses CalEDAS in all laboratory aspects of its bioassessment program (from sample log tracking to data analysis) and is currently updating the database with older datasets produced in MS Excel spreadsheets. Although the DFG does not provide technical support for this database, the ABL is willing to share working copies of the database in its current form with other laboratories.

3.4.3 Standardization of Reference Site Selection for Wadeable Streams

Variation is fundamental to biological communities and measures of biotic integrity based on these communities vary accordingly. Most bioassessment techniques account for variation through the use of reference sites. Since practical considerations limit our ability to find "undisturbed" or even "minimally disturbed" sites, most reference condition approaches seek to identify a compromise, the "least disturbed condition". Once candidate reference reaches have been identified, these can be used to characterize the range of biotic conditions expected for minimally disturbed sites.

For both the Russian River and San Diego IBI, the relatively subjective technique of "best professional judgment" (BPJ) and some semi-quantitative selection criteria were used for

selecting reference sites. These early studies have demonstrated the need for a framework for interpreting community data that can be applied in a standardized manner throughout the state.

At the February 2001 Western EMAP Reference Condition workshop in Phoenix, AZ, the workgroup drafted an approach to identifying reference sites that provides a strong framework for standardizing reference site methodologies. In May 2000, the DFG and Dr. David Herbst of SNARL collaborated to develop a quantitative approach to selecting reference sites in California. The basic approach uses landscape analysis tools (i.e., Geographic Information Systems, GIS) to identify areas within the region of interest that have minimal impacts (target areas). Field reconnaissance is then used to identify suitable stream reaches within these target areas, resulting in a pool of reference sites for the region of interest. The procedure consists of the following five steps:

1. Preliminary Organization and Prioritization
   a. Identify the region of interest and classes of streams to be evaluated
   b. Develop a list of land use disturbances of interest

2. Use GIS to Select Areas with Minimal Impact
   a. Divide the region of interest into areas that will serve as the basic reporting units of GIS analysis
   b. Summarize potential land use impacts for each area
   c. Determine impact scores using statistical properties of their distributions
   d. Use impact scores to identify regions with minimal disturbance: target areas

3. Ground Truthing
   a. Stage I– rapid reconnaissance.
   b. Stage II–identify ownership and obtain access permission.
   c. Stage III–intensive habitat scoring and selection of reference sites for sampling.

4. Sampling of Biotic Communities
   a. Sample a subset of the pool of reference sites for benthic invertebrates and analyze the data to define the range of biological metric values in the pool of reference sites.
   b. Reference sites may be sampled for other measures of stream or riparian health (e.g. fish/algal communities, water column chemistry, toxicity, etc.)

5. Iterative Refinement of the Reference Pool
   a. Refine the reference site pool based on biological, chemical and physical habitat data collected at each site.
   b. Eliminate or add candidate reference sites as land use changes occur.

This quantitative approach to selecting reference sites will be used by SNARL for developing an IBI in the eastern Sierras for the Lahontan Regional Board and by ABL for all other regions of California. For all past projects, where BPJ was used to select reference sites, this approach will be applied to assess the accuracy of BPJ selections. Currently, the ABL is using this quantitative approach for selecting reference sites in the Sierra Nevada Foothills Ecoregion and Central

Valley streams for the Central Valley Regional Board and the Sacramento River Watershed Program.

3.4.4 Development of an Index of Biological Integrity (IBI) for California

While there are many potential methods for evaluating biotic condition from community data, most approaches in the United States use a combination of multimetric and multivariate techniques. In multimetric techniques, a set of biological measurements ("metrics"), each representing a different aspect of the community data, is calculated for each site. An overall site score is calculated as the sum of individual metric scores. Sites are then ranked according to their scores and classified into groups with "good", "fair" and "poor" water quality. This system of scoring and ranking sites is referred to as an Index of Biotic Integrity (IBI) and is the end point of a multi-metric analytical approach recommended by the EPA for development of biocriteria (Davis and Simon 1995). The original IBI was created for assessment of fish communities (Karr 1981), but was subsequently adapted for BMI communities (Kearns and Karr 1994).

The first demonstration of a California regional IBI was applied to the Russian River watershed in 1999 (Harrington 1999). The Russian River watershed drains the third largest area in California, sustains an important anadromous salmonid population and is subject to a wide range of land uses including a variety of agricultural, timbering and urban development land uses. This demonstration IBI was based on a conceptual model described by the U.S. Environmental Protection Agency for development of numeric biocriteria. Benthic macroinvertebrates (BMI) were collected from 35 reaches within 21 tributary streams and the main stem of the Russian River during the fall 1995 and spring 1996 and 1997 using the CSBP. Although there was no indication of strong seasonal variability in the BMI communities, it was recommended that the index period for the Russian River tributary streams be in the spring. Since the original IBI was developed, samples have been collected annually (1998-2001) from the original sites and some additional locations.

As the Russian River IBI was being developed, DFG began a much larger project for the San Diego Regional Board. After a pilot project conducted on the San Diego River in 1995 and 1996, the San Diego Regional Board contracted DFG to help them incorporate bioassessment into their ambient water quality monitoring program. The initial sampling strategy was designed to gather a baseline of information to support several project goals:

- To include biological information in the San Diego RWQCB's ongoing water quality monitoring programs
- To create a species list of BMIs known from the region
- To establish a biological classification of different stream types in the region
- To identify potential reference sites for the San Diego regional bioassessments
- To determine the best index period for sampling BMI communities
- To select appropriate metrics for southern California stream bioassessments

During 1997 through 2000, data was collected from 93 locations distributed throughout the San Diego region. Most of the initial sampling sites were chosen to supplement chemical data

collected from long-term sampling locations, but some were established as reference sites based on "best professional judgment". In 2001, a new set of sites were chosen and sampled to further establish reference conditions in the San Diego region. The results of this sampling event were combined with the results of earlier sampling events to establish a preliminary IBI for the San Diego region. In July 2002, a final report was presented as a working IBI for the San Diego region.

Data from several sites sampled for the Los Angeles Regional Board were applied to the San Diego IBI with promising results. With additional refinement, the IBI developed for the San Diego region might be appropriately applied to all Southern California and perhaps Central Coastal wadeable streams and rivers. In 2002 and 2003, testing of impaired and potential reference steams will be conducted on data sets developed throughout this region using the CSBP.

The framework for developing an IBI for the Sierra Nevada Foothills Ecoregion and Central Valley streams will be available in 2004 and 2005, respectively. An IBI for wadeable coastal streams in northern California is being developed for the North Coast Regional Board. This IBI should be available in 2004 and will incorporate sites from the Russian River IBI that comply with the new quantitative approach to selecting reference sites, in addition to new sites throughout the region. Since this region extends from the Oregon border to south of San Francisco Bay, sites chosen by the San Francisco Regional Board will be tested and perhaps incorporated into Northern California Coastal IBI.